LEVEL

Research Memorandum 79-9

THE COMPUTER-ASSISTED MAP MANEUVER SYSTEM: A PRELIMINARY EXAMINATION OF ITS TRAINING EFFECTIVENESS AND SUITABILITY FOR USE AS A RESEARCH VEHICLE

AD A077981

Herbert F. Barber, John F. McGrew,
Steven R. Stewart and Robert S. Andrews

FORT LEAVENWORTH FIELD UNIT

DDC FILE COPY

DDC
RECEIVED
DEC 12 1979
D

U. S. Army
Research Institute for the Behavioral and Social Sciences

June 1979

79 22 5 113

Approved for public release; distribution unlimited.

(14) ARI-RM-79-9

Army Project Number
(16) 2Q163743A773

Command & Control
Training

(9) Research Memorandum 79-9

(6) THE COMPUTER-ASSISTED MAP MANEUVER SYSTEM: A PRELIMINARY EXAMINATION OF ITS TRAINING EFFECTIVENESS AND SUITABILITY FOR USE AS A RESEARCH VEHICLE.

(10) Herbert F. Barber, John F. McGrew,
Steven R. Stewart and Robert S. Andrews

Submitted by:
Robert S. Andrews, Chief
FORT LEAVENWORTH FIELD UNIT

(11) June 1979

(12) 97

Accession For	
NTIS GRA&I	X
DDC TAB	
Unannounced	
Justification	
By	
Distribution/	
Availability Codes	
Dist	Avail and/or special
A	

Approved by:

Abraham H. Birnbaum, Acting Director
Organizations and Systems
Research Laboratory

Joseph Zeidner, Technical Director
U.S. Army Research Institute for
the Behavioral and Social Sciences

Research Memorandums are informal reports on technical research problems. Limited distribution is made, primarily to personnel engaged in research for the Army Research Institute.

408 010

THE COMPUTER-ASSISTED MAP MANEUVER SYSTEM: A PRELIMINARY EXAMINATION OF ITS TRAINING EFFECTIVENESS AND SUITABILITY FOR USE AS A RESEARCH VEHICLE

CONTENTS

LIST OF TABLES

CONTENTS (Continued)

LIST OF FIGURES

INTRODUCTION

The training of the individual soldier as well as collective training has, in recent years, become a more challenging and difficult task. The dwindling resources, the competing demands for time and the more complex tasks to be trained that are presently being experienced by Army personnel increases the need to develop the most efficient and effective methods of training possible. In response to these demands and requirements, the Army Training Study (ARTS) was initiated to explore means of developing an "efficient, justifiable, and achievable training system for the Army of the mid-1980's,"[1] ARTS' approach to the problem was to formulate both a short term and long term effort. The short term effort, the Training Effectiveness Analysis (TEA) 78, was designed to capitalize on selected presently on-going training development and evaluation programs as a means to economically glean as much early information as possible and for deriving insights and direction for the longer term TEA 85 effort. TEA 85 is aimed at quantification of current training systems in order to support allocations of required training resources and as a basis for evolving training system improvements designed to provide the Army with the most efficient and effective training systems.

The Combined Arms Center (CAC) as proponent for all TRADOC battle simulations is presently involved in the development of several collective training programs for command groups at battalion and above levels.

---

[1] DA Msg, ATCG-ATS, DTG 221832Z Nov 77, Subj: Army Training Study (ARTS).

At the request of ARTS, these programs were examined to ascertain if any "piggy-backing" were possible to exploit this unique source of command group training data. There were two major areas of continuing concern in which an integration between the developmental effort for training command groups and the ARTS effort appeared feasible:

1. What is the relationship among command group performance as assessed in battle simulation, unit readiness, and combat effectiveness measures?

2. What are appropriate strategies to achieve optimum (cost and training effectiveness) command group proficiency levels through use of simulation technology?

Obviously definitive answers to these questions would not be possible for ARTS 78 within the time, resource and technology constraints prevailing. However, the need for and mutual interest in even tentative answers to portions of the above two questions prompted the initiation of an exploratory effort utilizing a current training system undergoing development by CAC: the Computer Assisted Map Maneuver System (CAMMS). It was determined that only five battalion command groups from two divisions were available for participation in this effort in time for input into the TEA 78 Report. It was recognized that the small sample size and limited representativeness would restrict generalizability and any analysis would therefore be of value mainly in terms of preliminary indications, trends and feasibility of approach.

However, it was felt that an initial look at command group training was necessary and should provide a source of useful planning information for ARTS as well as CAC.

OBJECTIVES

Specific objectives of this command group training effort derive from the broad questions identified above and the problem of measurement inherent in achieving answers to those questions. Only those objectives of direct relevance to the ARTS which could be addressed within the time constraints of TEA 78 are included below. Other objectives and analyses will be the subject of subsequent separate documentation.

Specifically, the present effort focused on three objectives:

1. To estimate the training effectiveness of CAMMS. This objective was aimed at the type and amount of performance improvement that can be produced with a battle simulation (i.e., CAMMS). Implicit in this objective is the development of command group performance assessment procedures and feedback mechanisms that would allow the command group to diagnose their training strengths and weaknesses and would enable meaningful feedback to the command group during and between training sessions so that their training effort could be concentrated in those areas where remedial training is most warranted.

2. To refine performance measurement procedures. This objective was designed to assist in the development of more valid and sensitive

means of measuring the proficiency level of battalion command groups. The command group/staff module of ARTEP 71-2 is a necessary though probably not sufficient component for such assessment. In addition, the degree to which measures for the performance of various of the tasks and subtasks could be made more objective and other measures of a more objective nature developed, the more uniform, equitable, and meaningful the assessment process and the greater the likelihood of achieving a basis for comparing performance across different battle simulations..

3. To estimate the feasibility of continuing to utilize CAMMS as a vehicle for investigating command group training. An important objective of this effort was to determine if CAMMS has the sensitivity and capability required of a research tool for providing sufficient data to answer some of the unanswered questions associated with command group training and to ascertain what might be necessary to enhance its utility for that purpose.

METHOD

PARTICIPANTS

Five battalion command groups, two mechanized infantry and three armor, participated in this effort. These groups were drawn from four brigades within two CONUS divisions.

SIMULATION SYSTEM

CAMMS,[2] as mentioned previously, is a training system undergoing development at the CAC. The system is being designed to overcome deficiencies of conventional methods, e.g., CPX, FTX, which have been used to provide command group training. It is a battle simulation designed to train commanders and staffs of armor, mechanized infantry, light infantry and cavalry units at both the brigade and battalion command levels. The command groups play within a non-nuclear environment and against a given enemy force.

CAMMS served as the instrument by which data on the performance of the abovementioned battalion command groups was obtained. A preliminary evaluation of the training effectiveness of the CAMMS was implicit in this process. For the purposes of this effort, the simulation was conceived to consist of three components. How each of these components was played or used in this effort, which parallels how the system is normally employed, is provided below:

---

[2]A more detailed description of CAMMS can be found in Battle Simulations and the ARTEP, CATRADA, Fort Leavenworth, Kansas, November 1977.

The Command Group

The command groups were constituted of those persons which normally would have been present under combat conditions. In addition to the commander and principal coordinating staff members, the groups included the Air Force liaison officer, the fire support officer and supporting NCO's and RTO's. The exercise was played within a simulated Tactical Operations Center (TOC) equipped with the type of communications gear normally issued to the battalion. Thus, the command groups had the capability to communicate with both their superior and subordinate units as well as adjacent units if such adjacent units were played.

Controllers

A number of controllers, whose primary purpose was to manage the exercise in such a way as to maximize the command group training experience, were used to conduct the exercise. This group included: (1) a chief controller who played the role of the brigade commander; (2) a brigade S1/S4 controller; (3) brigade S2 and brigade S3 controllers; (4) three maneuver company commander controllers; (5) a fire support controller and two to three supporting forward observers; (6) a direct air support controller; and (7) an opposing force (OPFOR) controller. All controllers, except those playing the brigade commander and principal staff, maneuvered "troops" and items of simulated equipment appropriate to their role on a game board that served as the terrain on which the battle was fought. Based upon the results of calculations produced by supporting computer

software and the events which were unfolding on the game board as the battle progressed, both the game board and brigade controllers provided realistic feedback and guidance to the command group players. An adjunct member of the controller team was a TOC monitor who observed the activities of the command group during planning and operational phases for the purposes of providing post exercise feedback and ratings of the groups' behavior for analysis in this investigation.

Computer Subsystem

The computer subsystem was designed to support military and logistics problems, greatly reduce map maneuver preparation time, provide faster and more accurate computations and, thereby, increase objectivity and provide a precise summary of the events which took place in the battle for analysis and critique purposes. The software available accommodated the employment of conventional forces with all their normal supporting weapons systems. Artillery, air, mortars, helicopters and admin/log functions were processed for the friendly force as well as the OPFOR. The system allowed for task organizations ranging from specific teams up to task force level. Interface with these programs was accomplished remotely through four computer terminals which were connected via commericial telephone lines to a centrally located computer. The terminal operators, normally radio-teletype operators, were provided by the participating units.

DESIGN

To explicitly address the first and implicitly address the second and third objectives of this effort, a pre-test/post-test design was used. Figure 1 depicts this design. The command groups participated in three separate exercises. However, the general scenario and type of mission were the same across the three exercises. The specific scenario[3] and mission used for each exercise was a variation of the more general one. Also, the three specific scenarios and missions were designed and assumed to be of equal difficulty. However, to correct for any differences in difficulty that may have existed, the missions for the pre- and post-test exercises were counterbalanced across units. A feedback session, the format and content of which were being pilot tested in this effort, followed each of the test and training exercises. Ideally, a control group which received only the pre- and post-tests would have been included to verify that those performance gains observed, if any, were due to the treatment or training rather than to other potential confounding variables. Additionally, it would have been desirable to have used CAMMS only for the treatment condition rather than across all conditions. This would have made it possible to avoid spurious relations that may be inherent in the results presented herein because the same measuring instrument was used both for testing and training purposes. Time and the availability of command groups precluded fulfilling either of these two conditions. These

[3]A brief description of the general and special situations for all exercises and the specific mission for each exercise is provided in Appendix A.

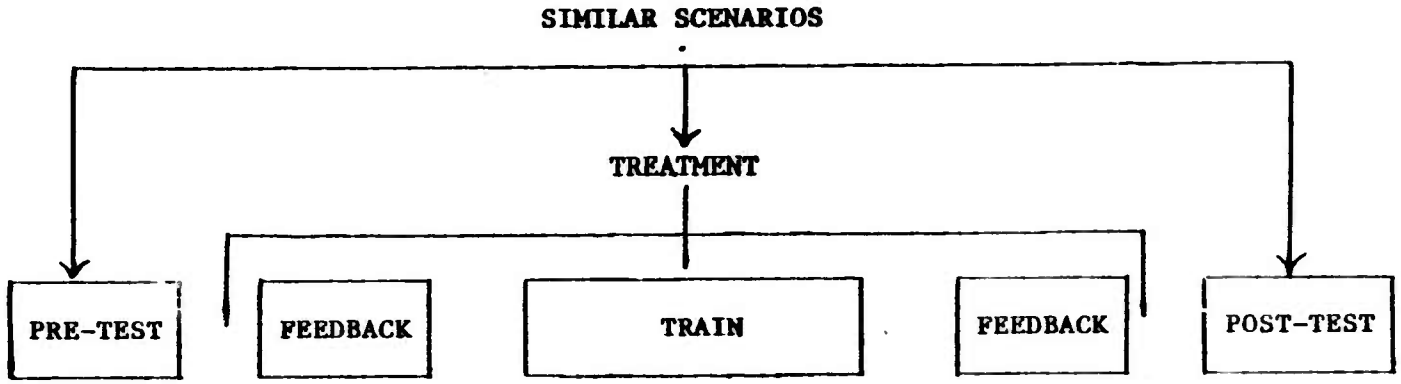

Figure 1. Experimental Design.

shortcomings as well as the size of the sample and instability within the controller group, a problem discussed more fully below, limit substantially the confidence which can be placed in the results presented in following sections.

BATTALION COMMAND GROUP PERFORMANCE MEASURES

A number of different types of performances were measured. These measures can be broken down into two broad categories; those which were obtained through a rating process and which, therefore, are subjective in nature and those which are objective in nature. Within both of these categories, performance measures related to a selected subset of the subtasks contained within the battalion/staff module of ARTEP 71-2[4] were gathered. Of the 61 subtasks included in the ARTEP, 47 were measured,[5] either subjectively, objectively or both. Appendix B identifies these 47 subtasks. Within the subjective category of measures were also included organizational process variables which have previously been used in the investigation of unit effectiveness. A more thorough discussion of the subjective and objective measures, in terms of the source from which they were obtained and the rationale for their investigation, is provided below.

---

[4]Army Training and Evaluation Program (ARTEP) for Mechanized Infantry/Tank Task Force, No. 71-2, Headquarters, Department of the Army, Washington, D.C., 17 June 1977.

[5]Performance of the remaining 14 subtasks was not measured either because these performances could not be observed or because they could not readily or realistically be elicited.

Subjective Measures

ARTEP Related. Ratings on each of the 47 subtasks were obtained from one or more members of the control team. The subtasks rated by each of the controllers who provided such information is summarized below:

Controller	Subtask Rated
Brigade S1/S4	3J, 3K, 9A, 9B, 9C, 9D
Brigade S2	1B, 2A, 2B, 2C, 2D, 3F, 3I, 5A, 5B, 5C, 5D, 10A, 12A
Brigade S3	1A, 1C, 1D, 1E, 1H, 3A, 3B, 3C, 3D, 3G, 3H, 4B, 6A, 6B, 7C, 8A, 8C, 8D, 10F, 10G, 11A
FSO	1I, 1J, 1K, 1L, 3E, 7A, 7B
TOC Observer	All subtasks, except 9A to 9D.

The controllers were instructed to base their ratings upon the conditions and standards stated in the ARTEP for each of the subtasks. Their ratings were provided on five-point Likert type scales with "forgotten, overlooked or unit failed to address this subtask" and "exceeded standard" serving as anchors on the low (1) and high (5) ends of the scale respectively. Two complete sets of ratings were obtained. One nearly complete set was provided by the TOC monitor and the second set was provided by the various controllers who responded to those subtasks which were within their area of staff responsibility. These two sets of ratings were averaged for analysis purposes.

In addition to ratings for each subtask, controllers provided overall estimates of how well the command group performed across all subtasks. Specifically, overall ratings of the performance of the S1, S2, S3, S4, FSO and ALO, battalion commander and command group as a whole were obtained

from the chief controller and TOC monitor. Again, a five-point scale with low and high anchors respectively being "not effective at all" and "extremely effective" was used. The TOC observer and the chief controller were the only sources from which these overall estimates could be obtained. Averages of their ratings provided the raw data for analysis.

Organizational Processes. Olmstead, et al. (1973) has found within a military context that the ability of an organization to cope effectively with its environment is in large part a function of how well it manages to perform certain task clusters or processes which Schein (1972) had identified in earlier theoretical work. These processes and the definition of them used in this effort are presented below:

- Acquiring Information. The process by which the command group acquired information about its external and internal environment.
- Communicating Information. The process of transmitting information that was acquired to those parts of the unit that could act upon it.
- Decision Making. The process of making decisions concerning actions to be taken as a result of the acquired information.
- Communicating Instructions and Orders. The process of transmitting decisions and decision related orders and instructions to those parts of the unit that must implement them.
- Monitoring. The process of obtaining follow-up information about the results of the plans and orders.

The process variables provide a source of data concerning command group behavior which should be related to the performance of clusters of ARTEP

subtasks. However, it was not known to what extent, if any, they would be related to performance of the subtasks and, thus, they constituted another potentially independent or nearly independent source of information concerning the battalion's performance. Although data are reported in subsequent sections on the process measures, time constraints precluded examining the relationships among subtasks and process variables. Such analyses will be performed, however, and included in subsequent separate documentation.

Objective Measures

ARTEP Related. Objective performance measures were developed for a subset of ARTEP subtasks whose standards were amenable to more rigorous quantification. These more objective measures were developed with the aid of military experts and through a process which involved decomposing a subtask into its more fundamental elements. For example, the development of a plan for the resupply of units in contact, one of the S4 ARTEP subtasks, consists of a number of discrete as opposed to continuous components, e.g., planning the refueling operation, calculating time-distance factors between supporting and supported units and establishing priority for the prepositioning of antiarmor ammunition. Once the S4 had developed the plan, a cognizant member of the controller staff could merely determine whether or not each element was represented. Even though the example provided represents a dicotomous measure in the most fundamental sense, selected controllers actually responded in one of four ways to the presence

or absence of target behaviors. These four responses were: "Yes, the behavior occurred;" "Yes, the behavior occurred but only with a specific probe(s) injected to elicit it;" "No, the behavior did not occur;" and "No, the behavior did not occur even with probing." The responses were assigned scores of 4, 3, 2, and 1 respectively, reflecting the assumed ordinal relation of the alternatives. Therefore, no assumptions were made as to the equality of intervals between the response alternatives on this set of discrete scales and since probing was left at the discretion of the controllers, it was not possible to control the number of probes injected in an attempt to elicit the appropriate behavior. The scores for each sub-element or component of the various subtasks were averaged to produce a more objective measure of subtask performance.

Nineteen of the 47 ARTEP subtasks identified in Appendix B could be more objectively quantified. These subtasks and the controllers who were responsible for their evaluation are as follows:

Controller	Subtasks
Brigade S2	1B, 2B, 2D, 3F, 5A, and 5D
Brigade S3	1A, 1C, 1D, 1E, 3D, 6A, 6B, and 8C
Brigade S1/S4	3J and 9D
Fire Support	1L, 7A, and 7B

As can be seen, each controller responded to a unique set of measures devised for the tasks shown above. In one instance, for subtask 1L, the subtask was broken down into three sub-subtasks, each of which was then decomposed into more fundamental elements. 1L sub-subtasks for which scores appear later in the report are: coordination with the FO;

informing the company commanders of the fire plan; and informing the supporting artillery of the fire plan. It was necessary to deviate from the four category discrete response format for three subtasks included in this category. These subtasks are 1E, 3A, and 3D.[6] The measures obtained for each of these subtasks are as follows:

1E. Select/control key terrain

- Number of key terrain features in the area of operation used by the battalion.
- The total number of key terrain features in the battalion's area of operation.
- The number of barriers, obstacles, and reinforced terrain used by the battalion.

3A. Determine critical place

- The grid coordinates of the location the command group identified as being the enemy's main thrust area.
- The grid coordinates of enemy avenues of approach identified by the command group.

3D. Select control measures

- The number of control measures used at each of the following type boundaries: check points, coordinating points, contact points, link-up points.

---

[6]These measures could have easily been included in the experimental category because they are more speculative than the other measures discussed here. However, since they were developed with specific subtasks in mind, they were included here.

Single measures were derived from the raw numerical data provided for each of three subtasks.

EXPERIMENTAL MEASURES

The measures subsumed under this rubric represent an attempt to quantify variables which prior research and deductive analyses have shown or suggested impact significantly on the effectiveness of the unit. Included in this category as well are speculative measures of battlefield outcomes, measures that should logically be related to the proficiency of the command group as reflected in those conventional parameters of their performance discussed above. These measures, as opposed to the command group performances investigated, do not command a concensus of opinion within the Army community vis-a-vis whether or not they are important or are critical. Therefore, one can only conjecture as to their suitability for performance assessment. Some of the measures, e.g., locus of control, are speculative enough in the context of the scenario used so as to make it impossible to state on an a priori basis whether more or less of the variable's presence is "good" or appropriate. Whether it is good or not is probably greatly dependent upon the particular scenario being played and the type of terrain on which the battle is fought. However, for other measures, e.g., the number of times the task force units become decisively engaged during the covering force operation, it was possible to deduce what should represent more superior performance, i.e., the less frequently the units become decisively engaged, the better one could judge their performance to be.

Because of the nature of the measures falling within this category, they were not included in analyses reported dealing with the training effectiveness of CAMMS. For the training effectiveness analyses, only the command group performance measures described in the previous section were examined. The experimental measures were only entered into correlational analysis to determine if performance of them was in some way related to the more conventional measures of command group proficiency.

As was the case for command group performance measures, the experimental measures can be more or less clearly divided into objective and subjective categories. The measures falling within these two broad categories and the method used to derive them are briefly discussed below.

Subjective Measures

Mission Accomplishment and Components Thereof. The chief controller and TOC monitor provided estimates as to the extent to which the command group was able to accomplish major tasks of the covering force mission,[7] the mission which was played for pre- and post-test and training phases for each battalion. Additionally, a global measure of mission accomplishment was obtained. A description of the type of performance measured is the following:

- Enemy Thrust. This is a measure of the degree to which the battalion command group was able to identify the major enemy thrust.

[7]Department of the Army FM 71-2, The Tank and Mechanized Infantry Battalion Task Force, 30 June 1977.

The controllers indicated whether the major thrust had been identified within 1, 1-3, 3-5 and 5 or more kilometers. For purposes of data analysis, these categories were scored 1, 2, 3, and 4 respectively and, thus, a lower score should indicate more superior performance.

- Decisive Engagement. This is simply the number of times the task force became decisively engaged. The ratings provided were divided into four categories; never, once, twice, and three or more times, which were assigned scores ranging from 1 through 4 respectively for analysis purposes.

- Task Force Losses. This is an estimate of the friendly losses which were incurred during the operation. One of five categories of loss was selected by the raters. These categories ranged from 0% to 50% in 10% increments through the fourth category and more than 50% losses constituted the fifth response category. This scale was assigned numbers 1 through 5, with five representing the greatest loss of forces.

- OPFOR Losses. This variable is the counterpart of the friendly force loss estimate. Therefore, the same rating scale was used by the two controller raters who provided this estimate.

- Mission Accomplishment. This constituted an overall estimate of the extent to which the battalion accomplished its assigned mission. The components of the task force mission described above should represent components of this overall estimate. A five-point rating scale was used with "not at all effective" and "extremely effective" constituting the anchors at the low and high ends respectively.

These measures constituted the only type of subjective measures examined within this more speculative category of variables. They logically should be related to more conventional staff and command performance parameters when examined across all groups.

Objective Measures

Locus of Control. Lessons learned from the 1967 and 1973 Mid-East Wars suggests that the extent to which control is centralized can significantly affect the performance of fighting units. The Israelis, as opposed to their enemies, greatly decentralized control of forces and other assets to their battalion commanders, which made it possible for them to capitalize on fleeting moments of opportunity on the battlefield. It created conditions which were conducive to improvisation and innovation that would have otherwise been impossible. Although the contribution of this variable to Israeli successes could have been purely a function of the nature of the requirements being faced and the terrain on which the battle was fought, it seemed reasonable to investigate its influence within the context of the present investigation.

To measure this variable, a matrix was developed with organizational level (echelon within the battalion to which control could be vested) and category of asset (that which could be controlled) forming the major axes. This matrix is shown below.

Matrix Used to Develop the Locus of Control Measures

	Squad	Platoon	Company	Staff	Battalion Commander
Squad					
Platoon					
Company					
Other Units (e.g., Engineer)					
Tanks					
Mortars					
TOWs					
Air Support					
Artillery					
Redeye					
Other Weapons (e.g., Attack Helicopter)					

The OPORD was examined by one member of the research team in order to fill in the matrix. In almost all instances, the organizational unit having control of given assets was identified in the Order. In those instances where it was not, control was assumed to reside at the organizational level to which it normally would have been assigned. The number of assets controlled by each organizational level was multiplied by an arbitrary weight (1 through 5 for squad through battalion commander respectively) and these products were averaged across all command levels

to produce a final score. Thus, the higher the final score, the more centralized was the control of assets within the battalion. Although this is a crude approximation of reality, it should be somewhat indicative of the locus of control within the unit.

Intelligence Operations and Reaction to Battlefield Contingencies. Barber and Kaplan (in press) in previous and ongoing work with another battlefield simulation undergoing development at CAC, the Combined Arms Tactical Training Simulator (CATTS), have found that of the conventional ARTEP command group performances examined, two general areas appear to be of especially great importance. These areas are intelligence processing[8] and the ability of the command group to perform all these activities implicit in rapidly shifting forces in response to battlefield contingencies in order to mass the forces at the appropriate time and place. These general areas have been found to be important for two reasons. Relative to the performance of other ARTEP related measures, these areas are the ones on which the command groups perform poorly and, yet, they are ones that, on a tentative basis, appear to capture the largest amount of variance in overall command group performance measures. Given these findings, it appeared that these two areas should be the foci of attempts to develop more quantitative measures of performance. Such measures might not be practical to gather for feedback purposes

[8]It was assumed that this process consists of two components: information acquisition and transmission, and the processing of raw information into intelligence. The measures discussed here address primarily the information acquisition and transmission component.

during a unit's normal play of the simulation, but they could prove to be useful for future research purposes in pursuit of the overall objectives of TEA 85. Accordingly, an attempt was made to identify quantitative measures which (1) judgmental analysis suggested should relate to command group performance and, (2) could be readily obtained during the play of the simulation. A number of such measures were identified, the preponderance of which were associated with the shift and concentrate forces performance area. Since the intelligence area and operations, of which concentration and shifting of combat forces is a large part, are so closely related, it was not always possible to clearly distinguish between the two in terms of the measures developed. Keeping this fact in mind, a description of the measures grouped into the intelligence and concentration of forces areas is presented below.

- Intelligence Processing. Two measures conjectured to be related to this general performance area were identified. The first had to do with communications within the battalion. Communications transmissions examined were those from (1) the company commanders requesting information from the battalion, (2) the company commanders providing information to the battalion, (3) the battalion to the company commanders providing information and, (4) the battalion to the company commanders requesting information. The command net was monitored for a fifteen minute period during each hour of the exercise. Messages transmitted and received were categorized into those predetermined categories identified above.

The number of transmissions of each type which occurred during the 15-minute interval sampled each hour of the pre- and post-test and training exercises were summed and averaged. Thus, the average number of transmissions of each type per hour by exercise were available for analysis.

The second measure conjectured to be related to the intelligence process was the ability of lower level task force units to "see" the battlefield. "Seeing" the battlefield is obviously in part dependent upon the extent to which that battlefield can be observed from assigned battle positions. Accordingly, the distance between each task force unit's battlefield position and man-made or natural obstacle or terrain feature which obscured line of sight was obtained. Connecting the points (grid quadrants) of the obstacles and terrain features formed a polygon whose area could be readily calculated. These calculations were made for each of the battle positions and the areas were summed and averaged. Thus, the score produced reflects how much clear viewing area was available to any given task force unit.

- Shift and Concentrate Forces. Five measures were developed which indirectly could reflect the ability of the command group to effectively perform this complex set of activities. These measures and the procedure used to derive them are briefly described below.

  - Distance between battle positions and supply points. The distance between each of the battle positions and the main and alternate supply points specified in the OPORD were obtained and averaged across

all battle positions. The resultant mean distance perhaps should reflect how well the units could react to contingencies on the battlefield since anticipation of such conditions should have been explicit in planning the battle and supply positions established. Thus, to a certain extent, intelligence preparation before the battle and fidelity of logistical (S1/S4) and operations (S3) coordination could be reflected in this measure as well.

•• Distance among battle positions. The distance among all battle positions as specified in the OPORD and established on the game board was obtained and averaged across battle positions. The mean obtained reflects the average distance between any given battle position and all other battle positions. Identification of the correct location of the enemy's main thrust should determine in large part how the friendly forces are initially positioned, whether in tighter mass or spread more loosely across the anticipated battle area(s). It should be reflective as well of a scheme of maneuver which anticipates that rapid shifts of combat power may be required as the battle progresses.

•• Number of battle positions. This measure should be highly related to the one above and could reflect anticipated exigencies that may develop. The measure was straightforwardly obtained by counting the number of battle positions psecified in the OPORD.

•• Distance of specified battle positions from each battle. Again, this measure was conjectured to reflect the ability to anticipate contingencies and develop compromises in terms of positioning of forces that takes into account the knowns and unknowns of the situation. To

calculate this measure, the distance between assigned battle positions for the various task force elements and the point at which the battle actually took place was obtained. These distances were then averaged for all task force elements.

•• Engagement range of all battles fought. This measure is probably a more indirect than a direct measure of command group performance. It should reflect the ability of the task force units to successfully implement the covering force mission as the parameters of it have been specified in the OPORD. If one assumes that such a mission carries with it the task of engaging the enemy at the maximum range possible, to attrit their forces to the greatest extent possible, while at the same time to minimize losses and expenditures of resources of the friendly forces, then the greater the range at which skirmishes occur within a given envelope should be indicative of relatively superior performance. A number of parameters are therefore likely to contribute to how this particular measure might behave and, in that sense, it is even more speculative than the others included in the general experimental variable category. This measure was obtained at that point in time when the friendly and OPFOR controllers mutually agreed that an engagement should take place. The distance between the opposing forces was obtained for each engagement of each task force unit and averaged across all engagements. Thus, the score produced represents the average distance between opposing forces for any given engagement.

Battlefield outcomes. Measures of this variety, if they can be produced with any degree of fidelity via simulation, should not necessarily for any given unit but should across a number of units bear some relationship to the performance capabilities of the command and control process. To explore this relationship, data were gathered on three crude battlefield outcome measures. These were number of friendly forces lost, number of friendly force weapons lost, and amount of ground lost[9] during the operation. The first two measures were derived from summaries of friendly unit weapons and personnel status summaries that were provided on an hourly basis by the computer. The losses calculated during each hour were summed over hours and averaged to produce the mean number of losses that occurred during any given hour. No differentiation as to type of weapons system or category of personnel was made in the calculations. Further, although the exercise SOP specified maximum limits for resupply action for given time periods, it was not possible to ascertain how rigidly the units adhered to these specifications. The amount of ground lost was obtained by merely measuring the distance between the front line traces at the beginning and end of each exercise.

All the various measures within this experimental category were gathered during the pretest, the training and the post test exercises.

---

[9]In a covering force mission, it is recognized that losing a specified or implied amount of ground within a specified amount of time is expected. However, given that the amount of ground to be lost was a constant, losing more or less than one should have is indicative of how well the unit performed.

However, for purposes of the correlational analyses presented in the next section, only the training exercise data were used because a more representative sample of the behaviors being measured could be obtained during this extended session.

Reliability of Subjective Measures

As discussed implicitly above, ratings were provided by multiple controllers for almost all of the subjective command group and experimental measures. The TOC monitor provided a comprehensive set of evaluations for the measures in question. The Brigade S2 and S3, the FSO, and chief controllers each rated subsets of those on which the TOC monitor provided evaluations. Thus, the paired ratings were comprised of the TOC monitor's evaluations in combination with one of the lattermentioned controllers for most measures. To obtain an estimate of the amount of agreement across the two sets of common ratings, they were intercorrelated separately for the battalions drawn from each of the two participating divisions.[10] Also, the data within a division was examined across all battalions and all exercises (pre-, training, and post) in which those units played. The results of this analysis are shown in Table 1. The correlations, with only three exceptions, are all statistically significant. However, in terms of their practical significance, there is not strong agreement among the raters, particularly the TOC monitor and the FSO.

[10]It was necessary to examine interrater reliability by division since the personnel, including the TOC monitor, who provided the data were different between the two divisions. This lack of uniformity in controller/rater staff is more fully discussed in the next section.

TABLE 1
Interrater Agreement for Command
Group Performance and Experimental Measures

Type of Measure	Division #1 (3 Battalions)	Division #2 (2 Battalions)
• ARTEP Subtasks		
•• TOC Monitor and S2	.33* (n = 92)[1]	.42* (n = 59)
•• TOC Monitor and S3	.19* (n = 142)	.28* (n = 103)
•• TOC Monitor and FSO	.16 (n = 45)	-.08 (n = 40)
• Organizational Processes		
•• TOC Monitor and Chief Controller	.37* (n = 53)	-.02 (n = 30)
• Overall Performance Measures		
•• TOC Monitor and Chief Controller	.39* (n = 49	.62* (n = 35)
• Mission Accomplishment		
•• TOC Monitor and Chief Controller	.61* (n = 49)	.43* (n = 35)

*Significant at .01 level or beyond

[1]The n's shown are the number of subtasks or items common for both raters during all three exercises across all participating battalions.

The average correlation across both divisions and all measures is .34 indicating on balance that only slightly less that 12% of the variance in the two sets of ratings is shared in common. This suggests the evaluations provided by the raters tended to be unique rather than common, a situation which may have resulted from (1) the insensitivity of the instruments used, (2) unique samples of the same behaviors being observed from distinct vantage points, (3) restriction of range in most of the ratings which decreases the likelihood of detecting the extent of rater reliability via conventional correlational techniques and (4) some combination of these or other factors. The question as to whether or not to combine the subjective evaluations of the various raters would have been rhetorical or nearly so if interrater reliability had been found to be relatively high (e.g., .70 or above). This question obviously became one of real concern when the reliabilities were found to be moderately low. A decision was made to combine the ratings, as implicitly indicated in previous portions of this measurement section. The decision was predicated upon the assumption that aggregation of the ratings would perhaps provide a better measure of reality than either of the unique aspects of it that the raters were presumed to have observed. Additionally, had the assumption that the raters differed because they were responding to unique aspects of the same behavior been fallacious, one would still have been thrust into a quandry as to which of the two ratings was the most veridical, a question that could not

have been addressed with the data collected in this preliminary investigation. The problem of low rater interreliability and the potential positive or negative impact of combining the ratings are additional considerations that must be taken to account with the other limiting factors recountered herein in interpreting the results of this effort.

PROCEDURE

Data Collection Team

A team, consisting of members provided by the (1) division of the participating units and (2) CAC and ARI, collected data on all command group performance and experimental measures. The brigade level controllers and game board players were provided by the participating unit's parent organization. The same individuals served in their respectively assigned roles, ones for which they had prior experience in all but a few isolated instances, for all exercises conducted within a particular division.

The TOC monitor was a lieutenant colonel provided from an element of CAC under whose auspices and direction the battalion command group ARTEP was developed. Thus, he was very familiar with staff procedure and operations at the battalion level. Although it was initially planned for the same individual to accompany the collection team to both participating divisions, for varying reasons this was not possible. The alternate TOC monitor, who served as a data collector during those exercises conducted at the second division, however, was assigned to the same element of CAC from which the first monitor came. Even though

these two individuals spent some time together in an attempt to form a unified frame of reference from which their ratings would be made, it is not possible to state definitively that this objective was achieved. This condition then obviated the possibility of examining the extent of agreement between the TOC monitor and the chief controller (who was the same individual across all exercises) for those instances where they provided comparable measures. Furthermore, initial plans called for a second observer who would have performed the TOC monitor functions in the combat trains area; a condition, which if fulfilled, would have provided a consistent frame of reference for the S1/S4 functional areas. Again, this requirement could not be fulfilled and, thus, it was impossible to collect data totally consistent across all units in the S1/S4 performance area.

The chief controller, one of the primary subjective data providers, also came from CAC and was assigned to the element responsible for developing CAMMS. Thus, he was very familiar with the simulation and had served as chief controller for at least 10 battalion level CAMMS exercises before participating in this effort. The data which he was responsible for providing and the data collection requirements of the TOC monitor and those brigade level controllers provided by the unit were specified in the previous section.

Two ARI personnel formed the final compouent of the data collection team. These individuals gathered the raw data from which the objective experimental measures were derived. They also monitored ongoing activities during the conduct of the exercise to insure that procedure was

being followed vis-a-vis experimental design and control measures and assisted any of the data collectors/raters who had difficulty responding to or understanding the data collection instruments.

Test and Training Exercises

The procedure followed to gather data related to the subjective and objective command group performance measures and the more speculative experimental measures is shown in Figure 2. The entire sequence of events, from orientation session through the post-test debriefing session, took place over a four-day period with no more than ten hours of game play occurring on any given day for the three exercises. Prior to initiating the first exercise, an orientation session was conducted for the players and controllers. During the first part of this session, both the controllers and players were present. They were briefed by ARI personnel and the chief controller concerning the overall objectives of the ARTS effort and what part the present investigation and they would play in accomplishing those objectives. Further, they were familiarized with the experimental design and the schedule of events which that design would entail as well as the general scenario that would be utilized in each of the exercises. Special requirements and constraints that would be required were discussed and the participants were told that a report summarizing the outcome of the exercises, but which would not identify the battalions in terms of their respective performances, would be provided to the division commander. The command group was provided copies of the training objectives for the exercises which were, in effect,

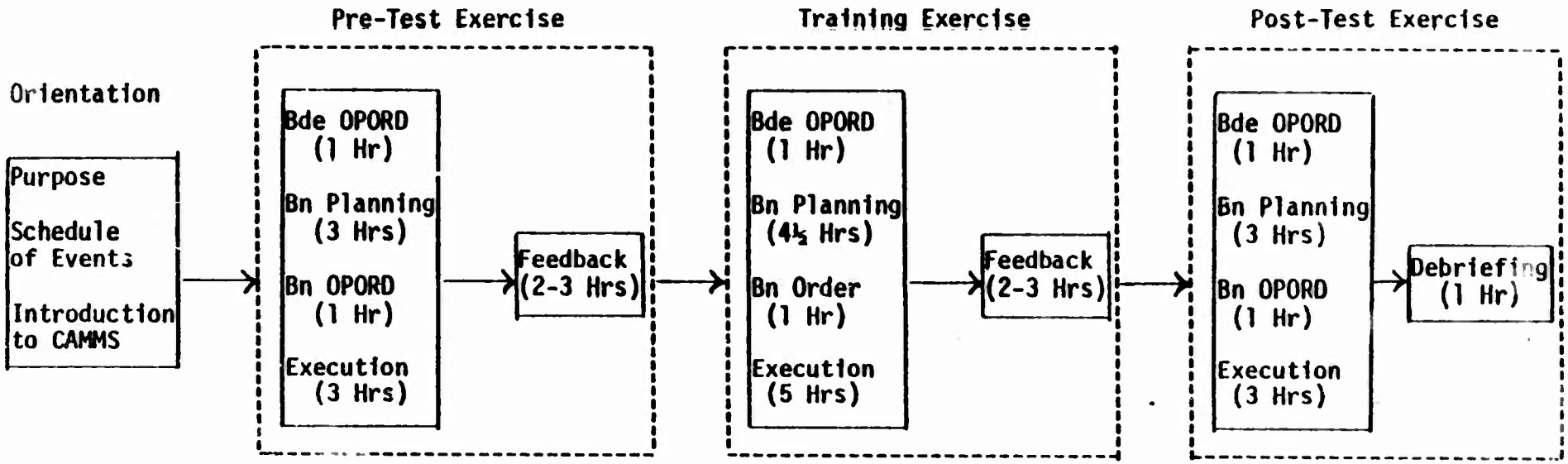

Figure 2. Experimental Procedure.

the 47 ARTEP subtasks on which most of the data were to be collected. During the second part of the session, only the controllers were present. They were (1) introduced to the data collection forms and provided instruction on how to use them, (2) provided guidance concerning more specialized procedural requirements than those which were covered in the general session, and (3) given training concerning how to perform the various roles and functions associated with the mechanics of the game as these factors related to the controllers' respective positions. A question and answer period was held at the end of the controller training session in order to verify that they fully understood the data collection requirements which they had been requested to fulfill and the requirements of the game which their assigned roles would entail.

Following the orientation session, the pre-test exercise was begun. During this time the chief and brigade level controllers issued the brigade OPORD to the battalion. The battalion command group spent three hours analyzing the order and developing their plans, a process which culminated in issuing their OPORD to the companies. The execution phase of the exercise began with the issuance of the battalion OPORD.

Following the execution, the data which had been collected throughout the planning and execution phases were consolidated and a feedback session was held. The general format of this session and the activities which immediately prefaced it are shown in Figure 3. A general observations briefing was held first with all members of the command group present. The chief and OPFOR controllers conducted this session during

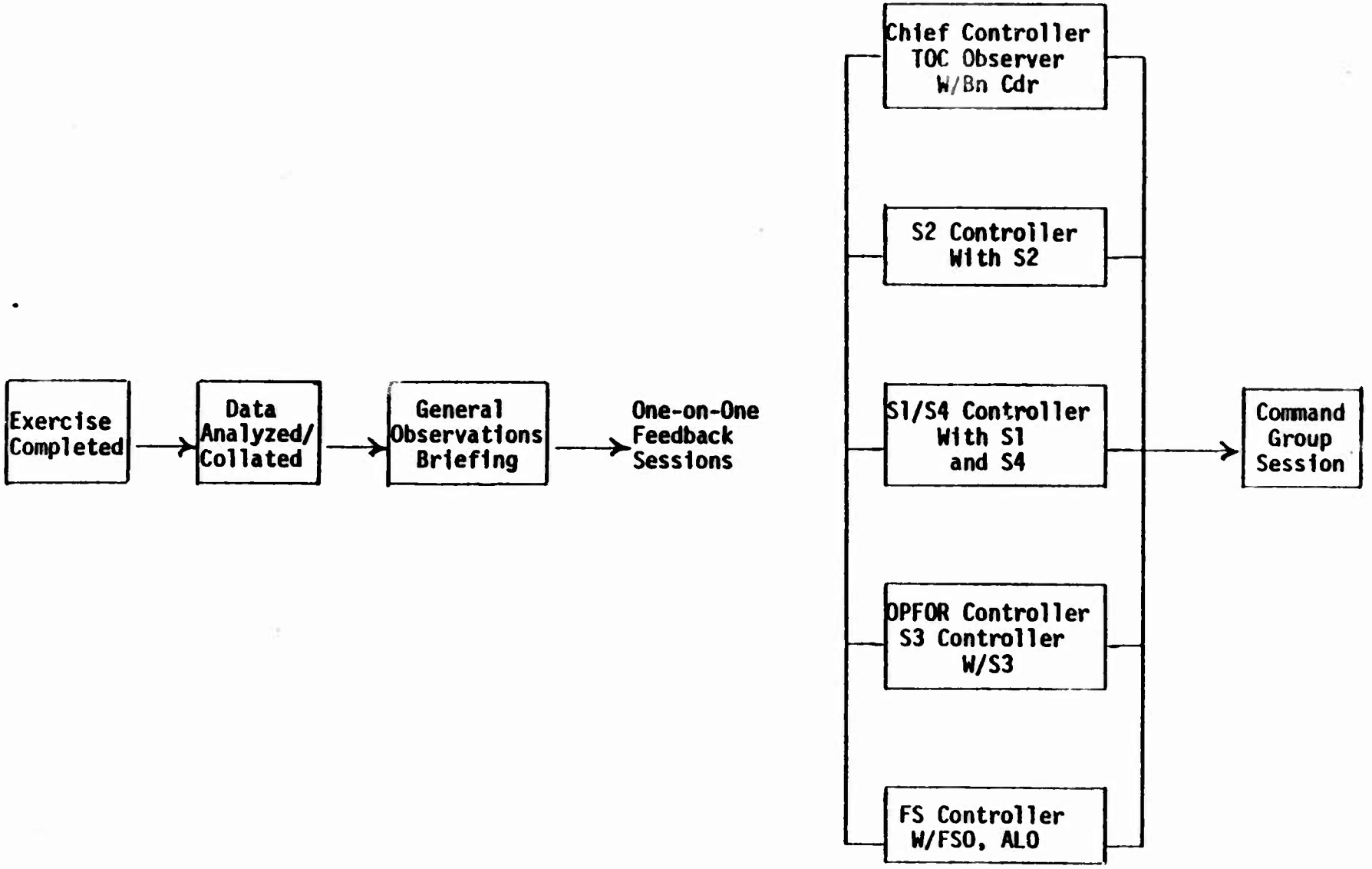

Figure 3. Feedback Session Format.

which they provided feedback based upon their observations and observations of others, as well as some of the quantitative data which had been collected. Following this briefing, each member of the command group and their counterparts on the controller staff met in separate closed sessions. During these sessions, detailed feedback was provided as to how well the staff member had performed those ARTEP subtasks which fell within his area of responsibility. The feedback was limited to the subjectively derived ARTEP subtask measures, the training objectives for the exercise. Specific examples of incorrect or inappropriate performance were provided for each subtask area to the extent this was possible. Following the one-on-one feedback sessions, the command group members met collectively in closed session during which each staff member had an opportunity to discuss the outcomes of the one-on-one session and propose to the battalion commander ways of overcoming deficiencies that had been identified. Remedial courses of action were developed and approved for implementation during the next exercise.

The sequence of events, from the issuance of the brigade OPORD through the feedback session, was followed for the training and post-test exercises with one exception. For the latter exercise, only a final general debriefing, rather than a full-blown feedback session, was conducted. In terms of duration, the execution phase was considerably lengthier in the training as opposed to the pre- and post-test exercises for obvious reasons. The exact duration of these phases is shown in Figure 2.

RESULTS

TRAINING EFFECTIVENESS

The primary basis used for assessing training effectiveness was the nature and amount of performance change occurring subsequent to the diagnostic (pre-test) exercise. Such data were tabulated for all appropriate dependent measures and repeated measures analysis of variance (ANOVAs) performed to determine whether any of the differences achieved statistical significance. For those performance dimensions that achieved statistical significance, a posteriori least significant difference tests were performed to identify which of the differences among means were significant, i.e., pre-test vs. post-test, pre-test vs. training and training vs. post-test. Results are presented and discussed for each major type of performance measure. Change in performances over time is not the only ingredient of training effectiveness of interest. Ability of measures to differentiate or distinguish strengths and weaknesses among the components of performance is important. Accordingly, the data were also interpreted from this perspective for each major type of performance measure. Because of time constraints and the large volume of pairwise comparisons which would be required to determine and present the statistical significance of the differences among the ARTEP tasks and subtasks, the degree to which these measures discriminated performance, as reported subsequently, is based on only visual inspection of the data.

Subjective ARTEP Subtask and Task Ratings

All main effects and interactions for both the exercise by subtask and exercise by task ANOVAs were statistically significant at the $P<.001$ level except for subtask effects which were significant at only the $P<.05$ level. Summary ANOVA tables are shown in Appendix C. Thus, there were differences in rated performance attributable to (1) the exercise session, (2) the subtask being judged and (3) the tasks comprised of those subtasks; and, of course, the interaction indicates that differences among exercise sessions were not uniform across all subtasks and tasks. The data showing the change in mean performance across the three exercise sessions for each subtask are shown in Table 2. A numerical improvement in mean performance between the first exercise session (pre-test) and the third (post-test) was obtained for 46 of the 47 subtasks. Though only 13 of these differences achieved statistical significance at $P<.05$ level, the consistency in the direction of the differences strongly suggests that the failure to achieve a far greater number of significant differences is more likely a function of the small number of units than it is the absence of real differences. These same data when aggregated by task produce findings which are consistent with those for subtasks as seen in Table 3 where differences between pre-test and post-test sessions are significant for only 3 of the 12 tasks but a numerical improvement in mean performance was obtained for all 12 tasks.

Examination of the differences in performance between the first (pre-test) and the second (training) session reveal patterns and magnitudes

TABLE 2

COMPARISON OF ARTEP SUBTASK RATING

FOR THE PRE, TRAIN AND POST TESTS EXERCISES

ARTEP Subtasks	Mean and Standard Deviation (S.D.) Values*						Difference Between Exercises		
	Pre (1)		Train (2)		Post (3)				
	Mean	S.D.	Mean	S.D.	Mean	S.D.	2-1	3-1	3-2
1A	4.20	.32	4.70	.27	4.50	.46	.50	.30	-.20
1B	3.40	.42	4.00	.29	4.10	.42	.60	.70*	.10
1C	4.00	.61	4.30	.27	4.50	.35	.30	.50	.20
1D	3.80	.27	4.00	.50	4.20	.27	.20	.40	.20
1E	3.80	.20	4.20	.45	4.60	.38	.40	.80*	.40
1H	4.20	.57	4.50	.35	4.60	.38	.30	.40	.10
1I	3.70	.23	3.80	.27	4.10	.13	.10	.40	.30
1J	3.70	.45	3.90	.22	4.20	.27	.20	.50	.30
1K	3.80	.41	3.80	.43	4.20	.27	0.00	.40	.40
1L	3.80	.27	3.90	.22	4.00	0.00	.10	.20	.10
2A	3.40	.42	4.00	.29	4.10	.42	.60	.70*	.10
2B	3.60	.42	4.20	.53	4.10	.26	.60	.50	-.10
2C	3.50	.61	4.20	.27	4.00	0.00	.70*	.50	-.20
2D	3.60	.42	3.20	1.81	4.00	0.00	-.40	.40	.80*
3A	4.10	.26	4.60	.42	4.40	.22	.50	.30	-.20
3B	3.90	.55	4.50	.50	4.40	.35	.60	.50	-.10
3C	4.20	.27	4.60	.42	4.50	.35	.40	.30	-.10
3D	3.70	.53	4.30	.45	4.50	.46	.60	.80*	.20
3E	3.80	.25	3.80	.27	4.10	.13	0.00	.30	.30
3F	1.40	1.34	3.70	.67	3.80	.54	2.30*	2.40*	.10
3G	3.60	.74	3.50	2.00	4.50	.35	-.10	.90*	1.00*
3H	4.00	.36	4.50	.36	4.50	.35	.50	.50	0.00
3I	2.10	1.34	3.30	.84	2.20	2.06	1.20*	.10	-1.10*
3J	3.90	.22	4.00	.09	4.00	0.00	.10	.10	0.00
3K	4.00	0.00	3.90	.22	4.00	0.00	-.10	0.00	.10
4B	4.10	.22	3.40	1.94	4.40	.42	-.70*	.30	.10

TABLE 2 (continued)

COMPARISON OF ARTEP SUBTASK RATING

FOR THE PRE, TRAIN AND POST TESTS EXERCISES

ARTEP Subtasks	Mean and Standard Deviation (S.D.) Values‡						Difference Between Exercises		
	Pre (1)		Train (2)		Post (3)				
	Mean	S.D.	Mean	S.D.	Mean	S.D	2-1	3-1	3-2
5A	3.80	.25	4.20	.27	4.00	0.00	.40	.20	-.20
5B	3.50	.46	4.00	.35	4.10	.26	.50	.60	.10
5C	3.60	.42	4.00	.35	4.10	.26	.40	.50	.10
5D	3.60	.42	3.90	.22	4.00	.43	.30	.40	.10
6A	3.90	.42	4.50	.35	4.00	.65	.60	.10	-.50
6B	3.10	1.72	4.30	.45	4.40	.42	1.20*	1.30*	.10
7A	3.90	.55	4.30	.27	4.30	.45	.40	.40	0.00
7B	3.70	.45	4.40	.42	4.50	.35	.70*	.80*	.10
7C	4.00	.35	4.30	.34	4.10	.55	.30	.10	-.20
8A	3.90	.22	4.30	.45	4.30	.57	.40	.40	0.00
8C	4.10	.26	4.50	.35	4.30	.56	.40	.20	-.20
8D	3.80	.27	4.20	.27	4.20	.23	.40	.40	0.00
9A	3.60	.52	3.90	.31	4.20	.45	.30	.60	.30
9B	2.40	1.52	3.50	.51	4.20	.45	1.10*	1.80*	.70*
9C	3.60	.89	3.00	1.73	4.00	0.00	-.60	.40	.10
9D	2.40	2.19	3.80	.43	4.20	.48	1.40*	1.80*	.40
10A	2.30	.84	3.20	.81	3.60	.50	.90*	1.30*	.40
10F	4.10	.56	4.30	.45	4.10	.55	.20	0.00	-.20
10G	2.40	2.19	4.10	.22	3.10	1.75	1.70*	.70*	-1.00*
11A	4.00	.35	4.20	.45	4.20	.23	.20	.20	0.00
12A	1.30	1.79	3.40	.47	4.00	.09	2.10*	2.70*	.60
Column Mean	3.54		4.02		4.14		.48	.60	.12

‡ N = 5

* Critical Mean Difference at the .05 Level = .67

TABLE 3

COMPARISON OF ARTEP TASKS FOR THE

PRE, TRAIN AND POST TESTS EXERCISES

ARTEP Subtasks	Mean and Standard Deviation (S.D.) Values‡						Difference Between Exercises		
	Pre (1)		Train (2)		Post (3)				
	Mean	S.D.	Mean	S.D.	Mean	S.D.	2-1	3-1	3-2
Develop Plan Based on Mission	3.84	.21	4.10	.21	4.28	.16	.26	.44	.18
Initiate Intel Prep on Battlefield	3.53	.24	4.09	.27	4.04	.15	.56*	.51	-.05
Prepare and Organize Battlefield	3.70	.20	4.12	.32	4.21	.12	.42	.51	.09
Troop Lead	4.10	.22	3.36	1.94	4.40	.42	-.74*	.30	1.04*
See the Battlefield During Battle	3.62	.28	4.02	.24	4.02	.19	.40	.40	0.00
Control and Coordinate Combat Operations	3.81	.32	4.40	.38	4.40	.45	.59*	.59*	0.00
Employ Fires and Other Combat Assets	3.87	.27	4.32	.18	4.30	.41	.45	.43	-.02
Concentrate and Shift Combat Powers	3.91	.22	4.33	.33	4.24	.43	.43	.34	-.09
Manage CSS Assets	3.52	.49	3.70	.28	4.14	.15	.18	.62*	.44*
Secure and Protect TF	3.42	.38	3.85	.32	3.91	.36	.43	.49	.06
Troop Lead During Battle	4.00	.35	4.20	.45	4.16	.23	.20	.16	-.04
React to Situation Requiring Special Actions	1.26	1.79	3.36	.47	3.96	.09	2.10*	2.70*	.60*
Column Mean	3.55		3.99		4.17		.44	.62*	.18

‡ N = 5

* Significant at the .05 level or higher.

very much like those of the differences between the first and third session. Eleven of the 47 subtasks and 4 of the 12 tasks reached the $P<.05$ level of significance. A numerical improvement in mean performance occurred for 41 of the subtasks and 11 of the 12 tasks.

The parallelism of findings thus far noted do not extend to differences in performance between the second (training) and third (post-test) sessions. Again, from Tables 2 and 3 it can be seen that differences were significant for only five of the subtasks though three tasks still reached the $P<.05$ level. Perhaps more revealing is the generally much smaller differences obtained and the fact no numerical improvement was obtained for 19 of the subtasks and six of the tasks. This apparent slowdown in the learning rate is suggestive of the common learning curve asymptote. This will be discussed along with other plausible explanations in a subsequent section of the report dealing with CAMMS as a training research vehicle.

In terms of differentiating among ARTEP subtasks, inspection of the subtask means within exercise session as shown in Table 2 reveals scores ranging from a low of 1.30 for one subtask to a high of 4.20 for three subtasks in the pre-test. Thirty-five of the subtasks received mean ratings below 4.00 which, based on the rating scale anchors used, indicate failure to meet the standard of performance. On the other hand, post-test session results show all but four subtasks performed at or above the standard. The range in means for that session extend from 3.10 to a high of 4.60. For diagnostic purposes, the data can be viewed

in terms of planning functions (Tasks 1 through 4) and execution functions (Tasks 5 through 12). From that perspective, the subtasks of identify critical combat information and intelligence (1B, 2A), gather critical information and intelligence (2B), analyze opposing force (2C), and disseminate critical combat information and intelligence (2D), develop a communications plan (3F) and plan/employ active/passive security measures (3I) were all subtasks observed during the planning stages that were rated relatively low (i.e., mean less than 4.0). During the execution phase, a similar pattern developed, i.e., subtasks dealing with the same general behaviors were rated relatively low. These low rated subtasks included gather critical information and intelligence (5B), analyze opposing force (5C), disseminate critical combat information and intelligence (5D), defeat or suppress opposing force's electromagnetic intelligence effort (10A), and react to opposing force electronic warfare (12A).

During the planning stages, those subtasks associated with the fire support area were also generally rated relatively lower. However, the fire support related subtasks during the execution phase did not follow this pattern.

During the execution phase, subtasks related to the admin/log area such as arm and fuel the systems (9A), fix the system (9B), and support the troops (9C), were relatively low rated. However, admin/log related subtasks were not rated particularly low during the planning phase. These results for the admin/log area may be less reliable than most of the

others because of the limited opportunity of the admin/log controllers, who were the only raters, to observe the performance within this particular area.

At the task level, and using the same criterion as in the subtasks, the results of comparing the mean ratings are basically the same as obtained by looking at the specific subtasks, but at a grosser level.

The results of the analysis of the performance of the battalion command groups on ARTEP tasks and subtasks indicate that three major areas of command group performance were the most problematic or poorly performed. These were: intelligence, fire support, and admin/log. These findings closely parallel the informal observations of the CAMMS Research Team. They also correspond to previous research on battalion command group ARTEP performance previously examined in CATTS exercises (Barber and Kaplan, in press).

Organizational Processes

Performance, as measured by ratings from 1 - not effective, to 5 - extremely effective, of the organizational process dimensions, was analyzed to determine whether there were changes occurring across sessions. The summary ANOVA table in Appendix D shows no statistically significant effects for exercise sessions, process dimensions or their interaction. Examination of the means and differences in Table 4 confirms that even on a nonstatistical numerical basis alone, the change across sessions is very slight albeit in the direction of improvement

TABLE 4

COMPARISON OF ORGANIZATIONAL PROCESSES

FOR THE PRE, TRAIN AND POST TESTS EXERCISES

Processes	Mean and Standard Deviation (S.D.) Values‡						Difference Between Exercises		
	Pre (1)		Train (2)		Post (3)				
	Mean	S.D.	Mean	S.D.	Mean	S.D	2-1	3-1	3-2
Acquiring information	3.20	.65	3.42	.55	3.50	.35	.22	.30	.08
Decision making	3.70	.27	3.90	.42	3.75	.79	.20	.50	-.15
Communicating instructions and orders	3.45	.27	3.65	.60	3.70	.48	.20	.25	.05
Communicating information	3.40	.38	3.45	.72	3.50	.40	.05	.10	.05
Monitoring	3.25	.25	3.85	.42	3.80	.45	.60	.55	-.05
Column Mean	3.40		.365		3.65		.25	.25	0.00

‡N = 5

from pre- to post-test and from pre-test to training. In terms of differentiation for diagnostic purposes, these data suggest that on the average, all five of these processes are performed a bit better than "somewhat effective" and that there is little basis for distinguishing among them. Whether this is attributable to the small sample, the insensitivity of the measures as used in this effort or other factors cannot be determined from these data. Further implications of these data will be discussed in subsequent section on measurement.

Overall Performance

The ANOVA for overall performance ratings obtained for each staff element, for the battalion commander and for the collective command group are contained in Appendix E. Exercise effects and overall command staff effects were statistically significant at the $P<.001$ level. When these data were analyzed in terms of simple effects, the only statistically significant differences among exercise sessions were for the personnel, logistics, and fire support functions. As shown in Table 5, all three differences were significant for personnel but only the difference between pre- and post-tests was significant for logistics and fire support. Like the ARTEP ratings, the differences in mean numerical values between pre-test and post-test for the six command staff elements individually and collectively, and for the battalion commander reflected improved performance as did the differences between the pre-test and training exercise. Again, a tendency toward asymptote

TABLE 5

COMPARISON OF OVERALL PERFORMANCE RATING

FOR THE PRE, TRAIN AND POST TESTS EXERCISES

In general, how effectively were the following activities performed?	Mean and Standard Deviation (S.D.) Values‡						Difference Between Exercises		
	Pre (1)		Train (2)		Post (3)				
	Mean	S.D.	Mean	S.D.	Mean	S.D.	2-1	3-1	3-2
Personnel (S1)	1.20	1.68	2.60	1.49	2.80	1.82	1.40*	1.60*	.20
Logistics (S4)	2.80	1.53	3.40	1.19	3.90	1.11	.60	1.10*	.50
Intelligence (S2)	3.10	.44	4.00	.60	3.90	.59	.90	.80	-.10
Operations (S3)	3.30	.32	3.90	.61	4.10	.81	.60	.80	.20
Fire Support (FSO)	3.00	0.00	3.90	.54	4.30	.46	.90	1.30	.40
Close Air Support (ALO)	3.70	.44	4.30	.31	4.30	.31	.60	.60	0.00
Overall, how effective was the BN CDR?	4.10	.52	4.30	.47	4.50	1.23	.20	.40	.20
Overall, how effective was the BN CMD GP?	3.40	.34	3.90	.49	4.10	.49	.50	.70	.20
Column Mean	3.13		3.79		3.99		.66	.86	.20

‡ N = 5

* Significant at the .05 level or higher.

for the change in performance ratings between the training and post-test exercises seems indicated.

Examining means by exercise across staff functions give evidence of differentiation. Scores range from a low of 1.20 to a high of 4.10 in the pre-test session. While the average rating for all but three of the functions is above the "very effective" level for the post-test, there is still a range of 2.80 to 4.50. Again, the personnel and logistics areas were rated relatively lower than the others.

ARTEP Related Objective Measures

The ANOVA tables for the five staff element/function categories for which objective measures for one or more subtasks were obtained are contained in Appendix F. In terms of training effectiveness as measured by differences among the three exercise sessions, only the exercise effects for the S1/S4 (admin/log) element achieved statistical significance at the $P<.05$ level. Table 6 shows that within this function, only one of the two subtasks produced significant differences. In this case, both the pre-test - post-test difference and the training - post-test difference achieved significance at the $P<.05$ level and both differences were in the direction of improved performance.

The degree to which the objectively rated component tasks of the various staff element functions produced differential performance scores is not appropriately addressed for these measures because they were not designed to be either representative or inclusive in number or content

TABLE 6

COMPARISON OF OBJECTIVE S1/S4 RATING

FOR THE PRE, TRAIN AND POST TESTS EXERCISES

Related ARTEP Subtask	Mean and Standard Deviation (S.D.) Values‡						Difference Between Exercises		
	Pre (1)		Train (2)		Post (3)				
	Mean	S.D.	Mean	S.D.	Mean	S.D.	2-1	3-1	3-2
3J	3.20	.62	3.40	.75	3.80	.49	.20	.60	.40
9D	1.90	.44	2.40	1.08	3.30	.58	.50	1.4*	.90*
Column Mean	2.55		2.90		3.55		.35	.10*	.65*

‡ N = 5

*. Significant at the .05 level or higher.

of the respective staff element functions. However, for purposes of completeness, the tables of means, S.D's, and differences for the S2, S3, and FSO elements are included in Appendix G. How they relate to the other ratings of the ARTEP subtasks from which they were derived will be treated in the section on measurement.

REFINEMENT OF PERFORMANCE MEASUREMENT PROCEDURES

The development of valid, reliable and sensitive measurement techniques to determine the proficiency level of battalion command groups was one of the objectives of this effort. As pointed out in the methods section, the measures used can be roughly dichotomized into those that reflect dimensions of performance that have been previously used and/or through concensus within the Army community, are generally accepted as being criteria for command group proficiency, e.g., ARTEP subtasks, and into those which are much more speculative in nature. The latter, though having a rationale and ostensibly a more objective dimensionality, are just being explored in regard to their relationship to proficiency. For these more speculative measures, the concern is whether and how these dimensions relate to more commonly accepted measures of command group performance and what difficulties there may be in their application and interpretation. Primary attention in this effort was on the latter type measures with the former serving as quasi criteria but both being examined in terms of relation to various measures of battlefield outcomes.

In the following paragraphs the interrelations among the various subjective and objective measures of battlefield outcomes are first

discussed. Next, the relationships of the battlefield outcome measures to both the subjective and experimental objective measures of command group performance are presented. Finally, the relation of the experimental objective measures to the subjectively assessed ARTEP subtasks is discussed.

Battlefield Outcomes

Table 7 shows the intercorrelations among the outcome measures. Five of the correlations are statistically significant and several others approach significance. The computer generated personnel and equipment losses correlate highly with each other suggesting that the algorithms presume attrition on these two dimensions to be closely linked. These same two dimensions plus area lost correlate highly and negatively with the factor enemy thrust, an estimate of the magnitude of error in identifying where the major penetration will occur. This suggests that friendly personnel, equipment and area losses will decrease as the error in estimating where the main thrust will occur increases. This paradoxical result may not be unreasonable for a covering force operation. It could mean that sufficient error in locating the enemy thrust will reduce the frequency and intensity of fire exchanges and thereby decrease personnel, equipment and area losses.

The substantial negative correlation between area lost and number of decisive engagements also seems plausible in that more frequent decisive engagements might for the short term reduce the area given up but are

TABLE 7
Intercorrelations Matrix of the Relationships Among the Various Battlefield Outcome Measures[a]

Battlefield Outcomes

	Subjective					Objective		
	Enemy Thrust (1)	Decisive Engagements (2)	TF Losses (3)	OPFOR Losses (4)	Mission Accomplishment (5)	Area (6)	Personnel (7)	Equipment (8)
(1) Enemy Thrust	1.00	.68	-.30	-.54	.65	-.87	-.94*	-.91*
(2) Decisive Engagements		1.00	.04	-.20	.10	-.89*	-.71	-.78
(3) TF Losses			1.00	.85	.87	-.11	.51	.48
(4) OPFOR Losses				1.00	.92*	.25	.75	.73
(5) Mission Accomplishment					1.00	.25	.76	.70
(6) Area						1.00	.80	.81
(7) Personnel							1.00	.99**
(8) Equipment								1.00

[a]Correlation based on n = 5

*Significant at .05 level, 2 tailed test

**Significant at .01 level, 2 tailed test

not likely to result in better mission accomplishment. The absence of a correlation between number of decisive engagements and mission accomplishment tends to confirm this reasoning.

It is reassuring to note that increases in the rated OPFOR losses correlate with improved missions accomplishment scores. Also, success in attriting the enemy is positively correlated with friendly force losses, a finding which is reasonable and to be expected.

Subjective Ratings and Battlefield Outcome Measures

Because of the experimental nature of the battlefield outcome measures, insights regarding their meaning for command group proficiency were sought by correlating these measures with the various ARTEP subtask ratings. The resulting correlation matrix with 376 entries contained 47 correlations statistically significant at $P<.05$. Five of the outcome measures accounted for 39 of these significant correlations with the remaining nine scattered across the other three outcome measures. Since any major import of the battlefield outcome measures in the present context resides in those 39 correlations, the tabled data and discussion are limited to those five outcome measures and the various ARTEP subtasks with which they correlate.

The first column in Table 8 contains the only negative correlations in the Table. For each of those four negative correlations, the interpretation would be that the smaller the error in identifying the location of the enemy thrust, the better the performance on the respective ARTEP subtasks. This seems to make sense given that three of the four ARTEP

tasks concern the planning and coordination of fire support and the fourth concerns integrating CSS into the scheme of maneuver. Plausible explanations for the positive correlation of "communicate/coordinate plans and orders" with amount of error in locating the enemy thrust are elusive. In the absence of other information, the correlation is as likely to be spurious as it is to be attributable to any of a number of remote explanations.

The relation of enemy losses to the seven ARTEP subtasks shown in the second column of Table 8 all seem reasonable. The better one analyzes the mission, determines critical place, organizes for combat, reinforces terrain, modifies scheme of maneuver, concentrates and shifts combat power, and fixes the systems, the greater the losses he is likely to inflict upon the enemy.

Mission accomplishment is the outcome measure having the greatest number of significantly related ARTEP subtasks. The three such subtasks relating to fire support planning again confirm the importance of this area. The fact that six subtasks concerned with preparing and organizing the battlefield are also highly related to mission accomplishment is not surprising. Similarly for the execution portion of the ARTEP, disseminating critical combat information and intelligence along with modifying scheme of maneuver, concentrating/shifting combat power and integrating CSS into scheme of maneuver were highly related to mission accomplishment as might be expected.

TABLE 8

Statistically Significant[a] Correlations with ARTEP Rating for the Five Battlefield Outcome Measures Having the Greatest Number of Significant Correlations

Subjective Outcome						Objective Outcome			
Enemy Thrust		Enemy Losses		Mission Accomplishment		Friendly Personnel Losses		Friendly Equipment Losses	
		1A	.88	1J	.87	1C	.85	1C	.85
1J	-.91	3A	.94	1K	.86	1J	.88	1J	.81
1K	-.85	3C	.94	1L	.87	1L	.88	1L	.81
1L	-.91	3H	.83	3A	.90	3A	.91	3A	.90
3G	.90	6A	.93	3B	.80	3C	.91	3C	.90
9D	-.85	8C	.93	3C	.90	5D	.81	9B	.81
		9B	.90	3G	.90	9D	.86	9D	.80
				3H	.90				
				3I	.80				
				5D	.93				
				6A	.97*				
				8C	.97*				
				9D	.92				

[a]All entries are significant at $P < .05$ level

*Significant at $P < .01$ level

Six of the seven ARTEP subtasks correlating significantly with computer generated friendly personnel losses also correlate significantly with computer generated friendly equipment losses. The previously mentioned high correlation between personnel and equipment losses was a precursor of this outcome but make the explanation of why better performance on those ARTEP subtasks were concomitants of greater personnel and equipment losses no less difficult. As proffered earlier, perhaps in a covering force mission the price typically paid for greater enemy attrition is increased friendly attrition.

Obviously the battlefield outcome measures may account for substantial variance in the performance of several other ARTEP subtasks which would not have achieved statistical significance with the present sample size. Correlations on the order of .57 to .70 between outcome measures and ARTEP subtask ratings would not achieve statistical significance even with a sample size of 10. Nonetheless, the data thus far are indicative of some underlying relationships which need to be further delineated in subsequent efforts and the relative importance or impact of the various subtasks on overall performance more fully explored. There is little reason to believe the various subtasks are equally crucial within any given mission. This notion was grossly examined with battlefield outcomes as the criteria by deriving three summary type ARTEP ratings and correlating them with the various outcome measures. These summary ratings were comprised by averaging those subtask ratings categorized under Tasks 1 through 4 for an overall planning phase score and those under Tasks 5

through 12 for an overall execution phase score. The planning and execution scores were again averaged to get a total ARTEP score.

Since there was no evidence available to the contrary, the subtasks were assumed to be roughly the same importance and hence, were weighted equally. Not surprisingly, the only correlations achieving statistical significance, as seen in Table 9, were between mission accomplishment and both the planning and total ARTEP ratings and even here only about 65% of the variance is accounted for. Harking back to the data relating the individual subtasks to the outcome measures, it is apparent that the summary data are reflecting primarily the influence of the 13 subtasks having a very high correlation with mission accomplishment. The relation of losses to the summary ARTEP ratings is negligible for the objective measures and though substantially higher numerically for the subjective estimates of loss parameters, is still short of statistical significance. Thus, it would seem that initial indications argue for differential weighting of the subtasks in any summary score.

Another indicator of overall command group performance was obtained with the organizational processes measure. Each separate process rating was included in this snalysis as well as the average rating across all five processes. Referring again to Table 9. there was a consistent and highly significant series of relationships between the organization process ratings and the subjective assessment of task force losses, OPFOR losses and mission accomplishment. The positive correlations

indicate that the higher the rating on the various processes, the higher the losses suffered by the opposing forces, the better the rating on mission accomplishment, and (perhaps as a result of the underlying relationship between OPFOR losses and task force losses) the higher the task force losses. Acquiring information seems to be about the only process dimension which did not have a high positive relation to mission accomplishment and OPFOR losses. Whether this is a function of inability to adequately observe and rate those activities, a bonafide lack of correspondence, or a spurious result attributable to the sample size is not known. Again, as previously noted for the ARTEP tasks, the objective outcomes did not correlate significantly or in a logically consistent manner.

The correlations of overall performance ratings with battlefield outcomes computed for each staff and separate coordinating element, the battalion commander and the command group as a whole are also presented in Table 9. Of the five significant correlations, two were for the S2 element - a positive correlation with number of decisive engagements and a negative correlation with area lost. Given the earlier mentioned high negative relationship between number of decisive engagements and area lost, this dual relationship is not surprising. It does not, however, make it any easier to understand how, in a covering force operation, the greater the number of decisive engagements, the better the rated performance of the S2 element. On the other hand, the high relationship of the close air support element rating to number of decisive engagements seems to

TABLE 9
Correlations Between Ratings of Command Group Performance and Battlefield Outcomes[a]

Ratings of Command Group Performance	Battlefield Outcomes: Subjective					Battlefield Outcomes: Objective		
	Enemy Thrust	Decisive Engagement	TF Losses	OPFOR Losses	Mission Accomplishment	Area	Personnel	Equipment
ARTEP Related								
Total	-.20	.43	.67	.75	.79*	-.18	.30	.22
Plan	-.28	.42	.69	.72	.83*	-.15	.33	.24
Execution	-.13	.42	.62	.75	.73	-.20	.20	.19
Organizational Processes								
Total	-.26	.26	.95**	.82*	.89*	-.21	.41	.35
Acquiring Information	.69	.35	.79*	.21	.47	-.72	-.32	-.27
Communicating Information	-.50	.04	.87**	.95**	.97**	.10	.65	.59
Decision Making	-.24	.13	.98***	.89*	.87*	-.16	-.47	.43
Communication Instructions	-.25	.44	.75	.71	.84*	-.21	.31	.22
Monitoring	-.12	.37	.91**	.81*	.83*	-.31	.30	.25

[a]Based on N = 5

*Significant at .05 level, 1 tailed test
**Significant at .01 level, 1 tailed test
***Significant at .001 level, 1 tailed test

TABLE 9 (Continued)
Correlations Between Ratings of Command Group Performance and Battlefield Outcomes[a]

Ratings of Command Group Performance	Battlefield Outcomes: Subjective: Enemy Thrust	Decisive Engagement	TF Losses	OPFOR Losses	Mission Accomplishment	Objective: Area	Personnel	Equipment
Overall Performance								
Personnel (S1)	.19	-.20	.57	.31	.13	-.24	.09	.17
Logistics (S4)	.65	.40	.26	-.22	-.23	-.74	-.53	-.49
Intelligence (S2)	.52	.89*	.21	-.20	.01	-.84*	-.59	-.66
Operations (S3)	.08	.72	.63	.38	.58	-.56	-.07	-.17
Fire Support (FS)	.16	.51	-.51	-.61	-.32	-.21	-.47	-.55
Close Air Support (CAS)	.46	.88*	-.25	-.29	-.15	-.58	-.60	-.69
Bn Cdr	-.67	.00	.76	.72	.93**	.21	.66	.58
Cmd Group	.00	.61	.75	.48	.67	-.50	.05	-.04

[a]Based on N = 5

*Significant at .05 level, 1 tailed test
**Significant at .01 level, 1 tailed test
***Significant at .001 level, 1 tailed test

make more sense, at least if viewed as a function of increased opportunity for effective performance. The final significant correlation is that between rated overall performance of the battalion commander and mission accomplishment. While this is a desired and reassuring outcome, it is not surprising given that the battalion commander is normally held responsible for the degree to which the assigned mission is accomplished.

Experimental Measures

The experimental objective measures of command group performance including the four categorizations of communications frequency were correlated with battlefield outcome indices which themselves are experimental in nature. The results are shown in Table 10. Trends are difficult to perceive in these data except in regard to locus of control and information provided to company commander variables. Significant and near significant correlations between these variables and enemy thrust location and personnel and equipment loss suggest the following picture. The higher (more centralized) the level of control, the greater the error in locating the enemy thrust with a consequent temporary reduction in loss of personnel and equipment for reasons discussed earlier. That reduction is concomitant with increasing amounts of information provided to the company by the battalion (perhaps as an attempt to compensate for perceived deficiencies in the performance of the company commander), thus, that information variable has a fairly high, though non-significant correlation with thrust location plus significantly high negative correlations with personnel and equipment loss.

TABLE 10

Correlations Between Experimental Measures of Command Group Performance and Battlefield Outcomes[a]

Experimental Measures of Command Group Performance	Battlefield Outcomes: Subjective: Enemy Thrust	Decisive Engagement	TF Losses	OPFOR Losses	Mission Accomplishment	Objective: Area	Personnel	Equipment
Locus of Control	.87	.60	-.60	-.59	-.73	-.62	-.88*	-.86
Engagement Range	-.27	-.73	.36	.37	.16	.39	.49	.59
BP from Battle	.38	.50	.70	.56	.41	-.63	-.10	-.11
BP from SP	-.35	-.02	-.41	-.46	-.11	.29	.01	-.06
Among BPs	.48	-.12	-.25	-.62	-.66	-.26	-.47	-.39
Clear Viewing	-.48	-.84	.23	.18	.13	.56	.56	.64
No of BPs	-.16	.15	-.63	-.35	-.20	.25	.11	-.18
No of SPs	-.19	.46	.14	-.34	.02	-.47	-.35	-.40
Info provided from Co Cdr	.70	.14	-.51	-.37	-.68	-.31	-.58	-.50
Info requested by Co Cdr	-.05	-.64	.19	.28	.03	.27	.31	-.42
Info provided to Co Cdr	.84	.67	-.67	-.73	-.76	-.63	-.93*	-.93*
Info Requested from Co Cdr	-.55	-.49	-.33	.21	.11	.75	.49	.50

[a]Based on N = 5

*Significant at .05 level, 2 tailed test

Further insights regarding the experimental measures were sought through correlation with the subjective ratings of performance on the 47 ARTEP subtasks. Fifty-eight of these correlations achieved statistical significance. Forty-four of these involved the same six experimental measures. To promote understanding and for the sake of brevity, only these 44 are presented here (Table 11).

All eight significant locus of control/ARTEP correlations were negative indicating that less centralized control results in better ARTEP subtask performance. Five of the six significant negative correlations for "battalion to company information" communication involve ARTEP subtasks which were also significantly related to locus of control. This can be interpreted to mean that less communication of information from battalion to company is associated with better performance, at least on those ARTEP subtasks where more autonomy is vested at lower levels. These subtasks concern priority of fires and fire support coordination, organizing for combat and reinforcing the terrain, disseminating critical combat information and intelligence, and integrating combat support services into the maneuver scheme.

All of the significant correlations between ARTEP subtasks and the "average distance between battle positions" measure were negative. This is an indication that at least for a covering force mission of the type, in the location, with the force structure, etc., as played and evaluated in this effort, better performance on certain ARTEP subtasks is associated

with smaller average distances between battle positions. Such a relationship seems quite rational when one considers that all but one of the ARTEP subtasks involved relate to the execution tasks of seeing the battlefield, control and coordination of combat operations, concentrating/shifting combat power, securing and protecting the task force, and the special action of reacting appropriately to enemy jamming.

One other experimental measure, "company to battalion information communication" produced all negative correlations with ARTEP subtasks. All but one of these related to planning activities. Three of the seven concern identification and gathering critical combat information and intelligence, and one to developing a communication plan. Why less communication of information up to battalion should be associated with better performance on those ARTEP subtasks is hard to fathom unless one assumes that the ratio of "noise" to "signal" increases with the amount of information communication from company to battalion. Though seven out of eight correlations of "company to battalion requests" with ARTEP subtasks were also negative, none of these subtasks were the same as those significantly related to the "company to battalion information communication" measures. In fact, most of the ARTEP subtasks significantly related to this variable involve fire support functions which seems to indicate that the less the company has to request information, the better these functions are performed. Further evidence of what has been previously discussed in regard to dissemination of critical combat information and intelligence is reflected in the negative correlation of performance on

TABLE 11

Statistically Significant[a] Correlations With ARTEP Ratings for the Six Experimental and Communications Measures Having the Greatest Number of Correlations

Levels of Control		Average Distance Between Battle Positions		Number of Battle Positions		Company to Battalion Requests		Company to Battalion Information		Battalion to Company Information	
1J	-.90	3H	-.87	1B	.90	1H	-.99*	1B	-.86	1J	-.83
1K	-.80	5A	-.96*	2A	.90	1J	-.87	1I	-.81	1L	-.83
1L	-.90	6A	-.81	2D	.99*	1K	-.87	2A	-.86	3C	-.91
3A	-.80	6B	-.93	3J	-.92	1L	-.87	2B	-.93	3H	-.91
3C	-.80	8A	-.93	3K	-.92	3I	-.88	3E	-.81	5D	-.89
3I	-.96*	8C	-.81	9A	-.92	5D	-.85	3F	-.85	9D	-.88
5D	-.91	8D	-.96			7B	.88	7C	-.89		
9D	-.92	10F	-.93			9D	-.87				
		12A	-.85								

[a]All entries are significant at $P < .05$ level

*Significant at $P < .01$ level

that subtask. Better performance on that subtask is associated with less frequent communication between company and battalion.

The final experimental measure for which significant correlations were tabled was number of battle positions. Three of these correlations are positive and three are negative. The three positive ones relate to identification and dissemination of critical combat information and intelligence, not an unreasonable outcome if your communications are adquate. The negative correlations are also reasonable in that they concern providing supplies and arming, fueling and maintaining the equipment. The greater the number of battle positions, the more difficult such tasks might be.

CAMMS AS A TRAINING RESEARCH VEHICLE

There are a number of observations, impressions and uncertainties generated from the conduct of this effort. Most result from the actual process of planning and running training exercises including discussions with cognizant player and controller participants. Hard data of a form amenable to statistical analysis are rarely available to support these findings although some derive from inadequacies identified as part of data analysis. The intrinsic value of these findings stems from the perceptions of the skilled CAMMS team and their intensive five-week exposure to and interaction with these exercises and the playing units. For purposes of exposition, these findings are roughly grouped into system factors and application factors. As a general prelude, it should

be stated that pervasive of both sets of factors is a well founded concern with the degree to which performance changes over time as assessed in CAMMS are a function of learning how to better perform as a command group as opposed to learning how to play CAMMS to get an improved score. This ambiguity could not be avoided within the constraints of the present effort nor would time and resources for TEA 78 permit a design from which these effects could be partially teased out. Essentially, the major need is for a yardstick or assessment tool external to CAMMS. It is intended that some preliminary data bearing on this question will be obtained if arrangements can be completed to have some of the units who play CAMMS, also play CATTS where any special skills unique to the CAMMS vehicle as opposed to content will be discounted.

Even with such additional data, the validity of the simulation and the measures of performance therein will require further confirmation. While considerable weapons, tactics and doctrinal experience, expertise and measurement sophistication have gone into developing CAMMS, a really sound assessment of the validity of CAMMS will be possible only after projects currently under development, e.g., Multiple Integrated Laser Evaluation System and the National Training Center are completed, thus producing a highly realistic "live" battlefield from which definitive measures of command group and troop performance can be obtained.

System Factors

Certain characteristics are requisite of a good training simulation. These include creation of a realistic environment within which events

unfold and actions transpire, reasonable representation of all functions and tasks which are deemed integral and important to the skills being trained, provisions to insure the performance of these tasks is precipitated, a means for credible assessment of the quality of performance and providing feedback regarding strengths and weaknesses of that performance, a capability to replicate on essential elements for equitable comparative assessments while maintaining sufficient flexibility to provide challenge and opportunity to make and correct mistakes. A training system possessing these characteristics has all the ingredients for supporting training effectiveness research. Obviously, these characteristics are not absolutes. However, all other things being equal, the better such requirements or characteristics are fulfilled, the greater should be the possibility of producing an optimally effective training system. Looking at CAMMS from this perspective, there are several observations which can be made. The most general and perhaps most important of which is that CAMMS is remarkably well along on most of these dimensions given its current stage of development. Other more specific observations include:

1. There is little basis for faulting realism. Though unquestionably greater fidelity could be achieved, there is no evidence of dissatisfaction on this dimension. In fact, CAMMS was judged superior to a CPX in a recent survey (Kaplan and Barber - in-press). Neither is there any evidence that greater fidelity would beget better training.

2. To the extent that the command group ARTEP is currently the guiding light for this sort of training, it may be desirable to include more extensive admin/log play. The relatively short duration of the execution phases of the respective exercises does not afford much opportunity for play and evaluation fo the gamut of these activities. Future modification of existing or generation of new scenarios should provide explicit attention to insuring that sufficient opportunity for play of all the staff element functions is provided. At the same time, consideration should be given to whether or not the ARTEP is sufficiently exhaustive. It is suspected that command group performance may be somehow more than the sum of the performances on the ARTEP tasks and subtasks.

3. A programmed or scripted set of proves exceeding substantially in number those developed for this effort will be necessary if behaviors of interest not normally forthcoming or not observably explicit in the play of CAMMS are to be elicited in some systematic fashion. This is particularly relevant for the planning phase where specific probes and/or situations will need to be developed to elicit the planning sequences so that they may be observed by the TOC observers and controllers. At the present time, most of the observation takes place during the delivery of the battalion OPORD and, consequently, very little information on the planning involved in its preparation is available.

4. Standards for the ARTEP subtasks are very generally worded and thus, can apply to almost any command group situation. This generality leaves much room for interpretation by the rater and consequently can

result in wide differences in ratings accorded the same group by different observers. Scoring or feedback based on such potentially divergent judgments is of questionable meaning and therefore, impact. Some elaboration of the standards for specific CAMMS training/evaluation exercises could help. The same may be said of the personnel and equipment loss data calculated by the computer. The current summary output must be re-analyzed to get the appropriate information for training feedback. The re-analysis of the output to produce both friendly and enemy personnel/equipment losses interferes with the time needed by the S1/S4 controller to perform his controller functions. Resupply is usually done for the friendly forces but not the enemy forces because the time to enter the resupplied items into the computer would force the exercise out of real time. In addition, the summary or end of exercise output makes no provisions to account for resupply in figuring the loss status of personnel/equipment. Some provision should be made either to increase the number of controller personnel to enable a timely re-analysis of the summary data or to devise a software package to analyze the summary data into a form that can be used for training feedback and to account for resupply.

5. Standardized elements for reliability of repeated administrations is an elusive target if you wish at the same time to allow for player innovation and challenging controller reactions, and to avoid a rote learning syndrome. For exportable training purposes as well as for training research/evaluation purposes where there must be dependence on

ad hoc controllers for many of the positions, a programmed text and accompanying "time into exercise" reference manual could provide additional assurances of consistency in use.

Application Factors

It is no less true with training systems than with weapons or other complex battlefield systems, implementation by and for people create opportunities for divergent procedures of use. Observations falling under the applications rubric in the current context are focused on those aspects of system use which either the trainer or trainee can capitalize upon, compensate for or at least be aware of in terms of possible impact on objectives.

It is often difficult in a home station environment to have available on a continuous basis for at least four days all principal members of a battalion command group. Leave, emergencies and other priority demands take their toll. While a surrogate in an individual position usually is not all that debilitating, it can impact on the overall performance of the command group and their learning rate. On the other hand, it can be argued that such turbulence is typical and therefore training and evaluation might be more realistic if this is allowed to occur. Subsequent analyses of the present data are planned in an effort to get a partial handle on the broader question of the effects of command group turbulence on performance insofar as it appeared as a variable across participating battalions. Hence, the concern is with length of tours with a unit etc., rather than presence or absence for a training exercise.

The concern with player substitution is paralleled by a concern by controller/evaluator substitutions occasionally within but more frequently between battalions making comparability tenuous because controller play cannot be totally proceduralized nor evaluation sufficiently objectified to insure that different controllers or evaluators reacting to the same events will behave the same. Consistency is largely dependent on having the same people from exercise to exercise within and battalion to battalion across player units. While failure to do so may not jeopardize the training value obtained, it can the training research benefits.

Providing the opportunity and data to support feedback to the playing unit is at least half the battle toward effective use. The remaining hurdle is the "how" or format for delivering that feedback. While only one feedback format was utilized, it appears that the technique was much more acceptable than some previously used. The command groups seemed receptive to the feedback and found it informative. The one-on-one feedback session appears to be in a format that is more comfortable than group feedback sessions for both the controller and the player within the command group. The controllers indicated that they felt much more comfortable and were much more willing to present negative feedback to the participants in the one-on-one session than in group settings.

One other pertinent aspect of application concerns the learning curve and the appropriate duration of CAMMS play to maximize performance improvement as a function of time. While there are many questions in this domain, the only evidence generated in the present study suggests that

performance has essentially asymptoted by the third exercise session (post-test). If that leveling off is attributable to a learning curve function, then more than two consecutive days play (at least of the same type mission) is not efficient. The prognosis might still be the same even if the result in attributable to an adaptation process where interest and motivation have decreased and standards have relaxed. Remedies for the latter condition might more easily be found. Actually the duration of a CAMMS exercise or successive exercises, the amount of time intervening between exercises, how these should be interspersed with use of other kinds of individual and collective training systems and exercises and many related questions are part of the overall training strategy the formulation of which will be dependent on achievement of various milestones in TEA 85.

CONCLUSIONS

The following conclusions are tempered to reflect the uncertainties resulting from constraints in the conduct of this effort, most notably, the small sample, and the potential confounding of measures of learning CAMMS with learning that which is trained by CAMMS:

- CAMMS shows evidence of being an effective training vehicle for improving battalion command group proviciency as subjectively judged in terms of differential performance on ARTEP Tasks and Subtasks and an overall assessment of the total command group and each of the major staff elements. The generally consistent and positive changes in performance

across pre-, training and post-test exercises and differentiation among subtasks and elements within exercise attest to its utility.

- Development of a greater number of objective measures of command group performance in CAMMS is feasible to both supplement and ultimately supplant some of the existing subjective ratings. It is a difficult and slow process and should not be expected to obviate the need for some subjective ratings.

- The relationship of command group performance to battlefield outcomes is complex. The quality of performance on no single measure yet identified can be adequately interpreted in isolation of other measures, or the condition of performance.

- Performance on some ARTEP subtasks appears to influence battlefield outcomes. Additional effort will be required to obtain reliable estimates of the degree to which performance on other ARTEP subtasks may also relate to outcome measures and what additional meaningful measures might be identified to more fully reflect the total dimensions of battlefield performance.

- Organizational process measures as used in this effort do not discriminate performance differences among the various measures themselves or change in performance as a function of the training exercise. However, their high relationship to mission accomplishment and other outcome measures warrant further investigations.

- CAMMS has the potential for fulfilling the requirements of a training and training research vehicle envisaged for pursuit of TEA 85

objectives. While some modifications are indicated, these are relatively modest in nature and several concern improvements which would be made in the normal evolution of CAMMS as resources permitted. The intrinsic worth and flexibility of CAMMS together with the data obtained and learning experienced on the part of the CAMMS team make it a leading contender for this role. Until other computer supported developmental simulations have progressed substantially beyond their current stage, CAMMS may be the only reasonable vehicle for examining the integration of troops on the ground with the play of a battalion level command group simulation.

REFERENCES

1. Barber, H. F. and Kaplan, I. T. Battalion Command Group Performance in Simulated Combat. ARI Technical Paper, in press.

2. Kaplan, I. T. and Barber, H. F. Evaluation of a Computer-Assisted Battle Simulation: CAMMS Versus a CPX. ARI Technical Paper, in press.

3. Olmstead, J. A., Christensen, H. E. and Lackey, L. L. Human Resources Organization Technical Report 73-19, August 1973.

4. Schein, E. H. Organizational Psychology (2nd Ed.), Prentice-Hall, Inc. Englewood Cliffs, New Jersy, 1972.

APPENDIXES

APPENDIX A

General Situation, Mission Assigned, and Assets Available to Participating Battalions

GENERAL SITUATION

Hostilities had broken out along the eastern German border approximately twenty-four hours before. The Park Land Forces had crossed the border and moved through the Fulda Gap area. Their movement had been progressively southwestward, with the intended objective conjectured to be Frankfurt. Since the enemy crossed the national border, their opposition had been a corps covering force.

The XX Armored Division had been ordered to establish a covering force forward of the FEBA and defend in sector in the main battle area. In turn, the division had given parallel missions to all three of its brigades. The brigade's specific mission was to establish a covering force along line Gold and delay in sector forward of the FEBA, defend in zone from the FEBA to the brigade rear boundary, retain city X, and be prepared to assist in the passage of lines of the corps covering force at line Delta.*

MISSION ASSIGNED AND ASSETS AVAILABLE

Each of the three brigade missions constituted the basis for establishing the mission for the pre- and post-test and training exercises.

*Note that line Gold, Delta, and the city to be retained were changed for the pre- and post-test and training exercises of the participating battalions.

Regardless of the specific brigade mission, the general mission given to the participating battalions was the same. The only difference was the terrain on which the operation was conducted. A description of the general mission of the battalion is the following: establish a covering force forward of the brigade main battle area, occupy and prepare various battle positions and battle areas, be prepared to delay from the various battle areas and battle positions, prepare bridges in sector for destruction, occupy given battle positions and areas in MBA upon completion of the covering force mission, and prepare to reinforce battle areas in MBA on order. To carry out this mission, each mechanized infantry battalion received an armor company in exchange for one of its companies and, similarly, armor battalions received a mechanized infantry company in exchange for one of their companies. Thus, the extent of task organization was controlled. In support of each battalion, regardless of type, was a platoon of combat engineers. Brigade assets accessible to the battalions included three battalions of 155mm SP howitzers, close air support, and attack and recon helicopters whose use was under the control of the brigade S3. Divisional GS artillery was also available through the brigade.

As pointed out above, the general mission of the participating battalions for all exercises was the same; only the terrain on which the mission was executed varied. However, the three battle areas or sectors were all in the Friedberg-Frankfurt general area and within 10 to 25 kilometers of one another. Within each area there were at least

two enemy avenues of approach, and fordable and non-fordable rivers to contend with. The terrain, in general, for all exercises was hilly woodland with flatter farmland interspersed throughout. For the training exercise, the terrain was less wooded and hilly and contained more farmland. There was also a greater number of avenues of approach available to the enemy.

BATTALION COMMAND GROUP ARTEP

1. Develop plan based on mission.
   - 1A. Analyze mission
   - 1B. Identify critical combat information and intelligence
   - 1C. Identify critical friendly information
   - 1D. Analyze friendly capabilities
   - 1E. Select/control key terrain
   - *1F. Select routes/zones to objective
   - *1G. Select battle positions
   - 1H. Select delay and covering force positions
   - 1I. Plan use of organic/attached and non-organic fires
   - 1J. Determine priority of fires
   - 1K. Determine fire support required
   - 1L. Conduct initial fire support coordination
2. Initiate intelligence preparation of the battlefield
   - 2A. Identify critical combat information and intelligence
   - 2B. Gather critical combat information and intelligence
   - 2C. Analyze opposing force
   - 2D. Disseminate critical combat information and intelligence
3. Prepare and organize the battlefield
   - 3A. Determine critical place
   - 3B. Select a course of action
   - 3C. Organize for combat
   - 3D. Select control measures
   - 3E. Plan organic, attached, and non-organic supporting fires and determine priority
   - 3F. Develop a communication plan
   - 3G. Communicate/coordinate plans and orders
   - 3H. Reinforce terrain
   - 3I. Plan/employ active/passive security
   - 3J. Provide supplies
   - 3K. Maintain equipment
4. Troop lead
   - *4A. Supervise preparations
   - 4B. Supervise compliance with TF order
   - *4C. Conduct rehearsals
5. See the battlefield during the battle
   - 5A. Identify critical combat information and intelligence
   - 5B. Gather critical combat information and intelligence
   - 5C. Analyze opposing force
   - 5D. Disseminate critical combat information and intelligence
6. Control and coordinate combat operations.
   - 6A. Modify scheme of maneuver
   - 6B. Coordinate/communicate changes
   - *6C. Supervise execution
   - *6D. Maintain the battlefield
7. Employ fires and other combat support assets
   - 7A. Modify fire support plan
   - 7B. Employ fires
   - 7C. Employ other combat support assets
8. Concentrate/shift combat power
   - 8A. Determine critical place and time
   - *8B. Concentrate/shift combat power in the attack
   - 8C. Concentrate/shift combat power in the defense or retrograde
   - 8D. Protect thinly held areas
9. Manage combat service support assets
   - 9A. Arm and fuel the systems
   - 9B. Fix the systems
   - 9C. Support the troops
   - 9D. Integrate CSS into scheme of maneuver
10. Secure and protect the TF
    - 10A. Defeat or suppress opposing force's electromagnetic intelligence effort
    - *10B. Defeat or suppress opposing force's imagery intelligence effort
    - *10C. Defeat or suppress opposing force's human intelligence effort
    - *10D. Deceive the opposing force
    - *10E. Reduce vulnerability to opposing force mass destruction weapons systems
    - 10F. Detect/impede threats to TF security
    - 10G. Detect/defeat opposing force air assets
11. Troop lead during battle
    - 11A. Supervise compliance with TF order
12. React to situations requiring special actions
    - 12A. React to opposing force electronic warfare
    - *12B. React to chemical or biological attack
    - *12C. React to nuclear attack
    - *12D. React to loss of key member of command group

*Not evaluated in CAMMS

THIS PAGE IS BEST QUALITY PRACTICABLE
FROM COPY FURNISHED TO DDC

REPEATED MEASURES ANALYSIS OF VARIANCE FOR ARTEP TASKS

Source	df	Mean Square	F
Exercise	2	6.17	20.20**
Error	8	.31	
ARTEP Tasks	11	1.97	6.83**
Error	44	.29	
E x ARTEP Tasks	22	.80	3.77**
Error	88	.21	

**Significant at the .001 level

REPEATED MEASURES ANALYSIS OF VARIANCE FOR

ARTEP SUBTASK

Source	df	Mean Square	F
Exercise	2	24.00	32.56**
Error	8	.74	
Subtask	46	2.67	4.99*
Error	184	.54	
E x S	92	.69	1.93**
Error	368	.36	

* Significant at the .05 Level
** Significant at the .001 Level

REPEATED MEASURES ANALYSIS OF VARIANCE FOR

ORGANIZATIONAL PROCESSES

Source	df	Mean Square	F
Exercise	2	.53	1.49
Error	8	.36	
Org Processes	4	.39	1.86
Error	16	.21	
E x OP	8	.07	.83
Error	32	.09	

REPEATED MEASURES ANALYSIS OF VARIANCE FOR OVERALL PERFORMANCE

Source	df	Mean Square	F
Exercise	2	7.91	8.87**
Error	8	.89	
Overall	7	6.08	4.53**
Error	28	1.34	
E x O	14	.22	.43
Error	56	.52	

**Significant at the .001 level

REPEATED MEASURES ANALYSIS OF VARIANCE FOR

OBJECTIVE S1/S4 RATING

Source	df	Mean Square	F
Exercise	2	2.57	8.70**
Error	8	.29	
S1/S4	1	6.91	67.22**
Error	4	.10	
E x S1/S4	2	.41	1.40
Error	8	.29	

**Significant at the .001 level

REPEATED MEASURES ANALYSIS OF VARIANCE FOR

S2 SCORES

Source	df	Mean Square	F
Exercise	2	.09	..16
Error	8	.58	
S2	5	.68	1.30
Error	20	.52	
E x S2	10	.40	1.08
Error	40	.37	

REPEATED MEASURES ANALYSIS OF VARIANCE FOR

S3 SCORES

Source	df	Mean Square	F
Exercise	2	7.01	.40
Error	8	17.73	
S3	8	67.24	1.85
Error	32	36.25	
E x S3	16	2.66	.30
Error	64	8.81	

REPEATED MEASURES ANALYSIS OF VARIANCE FOR

OBJECTIVE FSO SCORE (1La, 1Lb, 1Lc)

Source	df	Mean Square	F
Exercise	2	.14	.02
Error	8	5.78	
FSO	3	.50	.35
Error	12	1.42	
E x F	6	1.87	1.38
Error	24	1.36	

REPEATED MEASURES ANALYSIS OF VARIANCE FOR

OBJECTIVE FSO SCORE 7B

	Source	df	Mean Square	F
7B	Use of fires	7	13510.33	1.72
	Error	6	7834.66	

COMPARISON OF THE OBJECTIVE S2 SCORES FOR THE PRE, TRAIN AND POST TESTS EXERCISES

Measurements	Mean and Standard Deviation (S.D.) Values‡						Difference Between Exercises		
	Pre (1)		Train (2)		Post (3)				
	Mean	S.D.	Mean	S.D.	Mean	S.D.	2-1	3-1	3-2
1B	3.33	.38	2.74	.41	3.12	.28	-.59	-.21	.38
2B	3.11	.72	2.82	.92	2.90	.24	-.29	-.21	.12
2D	3.45	.32	3.48	.24	3.32	.20	.03	-.13	-.16
3F	2.85	.85	3.25	.82	2.62	.91	.40	-.23	-.63
5A	3.12	.55	3.46	.64	3.50	.39	.34	.38	.04
5D	3.22	.54	2.66	.53	3.42	.81	-.56	.20	.76
Column Mean	3.18		3.07		3.15		-.11	-.03	.08

‡ N = 5

COMPARISON OF S3 SCORES

FOR THE PRE, TRAIN AND POST TESTS EXERCISES

Measurements	Mean and Standard Deviation (S.D.) Values‡						Difference Between Exercises		
	Pre (1)		Train (2)		Post (3)				
	Mean	S.D.	Mean	S.D.	Mean	S.D.	2-1	3-1	3-2
1A	3.4	.89	3.0	1.73	3.0	1.79	-.4	-.4	0
1C	4.0	0	3.2	1.79	3.2	1.79	-.8	-.8	0
1D	3.7	.30	2.86	1.66	2.96	1.68	-.84	-.74	.1
1E	7.7	4.27	10.16	18.67	6.46	9.72	2.46	-1.24	-3.7
3D	2.06	1.13	2.36	1.97	2.3	1.79	.3	.24	-.06
6A, 8C	3.0	1.70	3.94	.40	1.68	2.12	.94	-1.32	-2.26
6B	1.88	1.99	2.10	1.64	1.84	1.82	.22	-.04	-.26
6B, 8C	.88	1.07	1.76	1.71	.88	1.97	.88	0	-.88
8D	1.14	2.28	.72	.96	.80	1.79	-.42	-.34	.08
Column Mean	3.08		3.34		2.57		.26	-.51	-.77

‡N = 5

COMPARISON OF OBJECTIVE FSO SCORES

FOR THE PRE, TRAIN AND POST TESTS EXERCISES

Measurements		Mean and Standard Deviation (S.D.) Values‡						Difference Between Exercises		
		Pre (1)		Train (2)		Post (3)				
		Mean	S.D.	Mean	S.D.	Mean	S.D.	2-1	3-1	3-2
1La	Coordination with FO	2.90	1.45	2.40	1.91	3.20	1.79	-.50	.30	.80
1Lb	Inform C.Co of fire plan	3.3	1.61	2.80	1.79	3.20	1.79	-.50	-.10	.40
1Lc	Inform Arty of fire plan	3.2	1.59	3.00	1.73	3.00	1.73	-.20	-.20	0.00
7A	Modify fire support plan	1.90	1.34	3.80	.43	2.50	1.66	1.90	.60	-1.30
7B	Use of indirect fires‡	58.5	18.1	153.5	137.13	164.0	74.79	95.0	105.5	10.5

‡N = 5

‡N = 4